Performance comparison of GES DISC data as a service between server-based system and Cloud Computing platform

IN33B-1537

Long Pham[1], Aijun Chen[1,2], Eric Winter[3], Christopher Lynnes[1]

[1]NASA Goddard Earth Sciences Data & Information Services Center (GES DISC); [2]Center for Spatial Information Science and Systems, George Mason University; [3]Wyle Information Systems, at GES DISC

Contact: Long Pham
long.b.pham@nasa.gov

Introduction

The NASA Goddard Earth Science Data and Information Service Center (GES DISC) recently demonstrated and evaluated the provision of (1) "Data-as-a-Service" in the Goddard Space Flight Center (GSFC) Cloud platform using OPeNDAP (Open-source Project for a Network Data Access Protocol) protocols; (2) "Data-processing-as-a-Service" using GrADS Data Server (GDS), and (3) "Data-analysis-as-a-Service" using NASA Giovanni (GES DISC Interactive Online Visualization ANd aNalysis Infrastructure). The latter was also evaluated on the commercial Cloud of Amazon Web Service (AWS).

The work entailed porting OPeNDAP to the NASA GSFC Cloud, along with a selected set of data (AIRS/Aqua, OMI/Aura, TRMM, and MERRA*), and then exercising the server using several clients from different network. The evaluation included the following aspects of using open source software in the Cloud to serve large volumes of satellite data for public access and subsetting.

- Ease of porting and operating OPeNDAP, GDS, and Giovanni in Cloud (e.g., time needed to setup one instance).
- Performance factors, including accessibility, stability, speed, and scalability, as compared with existing GES DISC capabilities.

*AIRS: Advanced Infrared Sounder; OMI: Ozone Monitoring Instrument
TRMM: Tropical Rainfall Measuring Mission; MERRA: Modern-Era Retrospective Analysis for Research and Applications

Ported Software: OPeNDAP, GDS, Giovanni

The following applications were ported on the NASA GSFC Cloud Platform and Amazon EC2 (Elastic Cloud Computing) platform.

A. OPeNDAP – Data Access-as-a-Service

OPeNDAP is an open source data service. It provides a way to easily share science data across the internet. It is widely used in earth-science research settings and supports multiple commonly used data format including HDF4/5, NetCDF3/4, CSV, and ASCII. It facilitates Earth Science data online sharing and access and makes data online access convenient to scientists. The OPeNDAP 1.7.0 was installed and tested on the NSAS GSFC Cloud and Amazon EC2 platform.

B. GrADS Data Server (GDS) -- (Data Processing-as-a-Service)

GDS is a data server providing stable and secure subsetting and analysis services across the Internet. The GDS 2.0 was installed and tested.

C. Giovanni -- (Data Analysis-as-a-Service)

Giovanni is an interactive Web-based application for visualizing, analyzing, and accessing vast amounts of Earth science satellite data. There are different Giovanni portals available for various kinds of satellite data, e.g. MODIS, AIRS, OMI, CloudSat, etc.

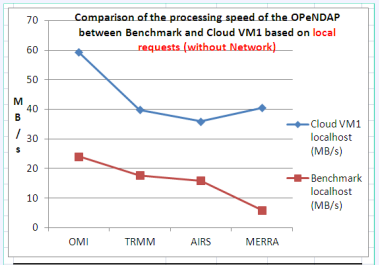
Test Data for OPeNDAP

Four satellite data products were selected as test data for OPeNDAP (25 granules each), as listed in the following table:

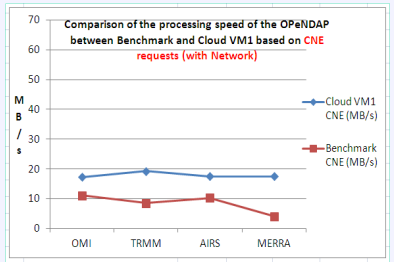
Data Products	Server Specification	Input Data Format/ Size (MB)/granule	Output Data Format/ Size (MB)/granule
OMI/Aura	Local: 16 cores * 2.4GHz 32GB RAM	HDF5/5.62	NetCDF/68.46
TRMM	Same as above	HDF/81.14	NetCDF/81.25
AIRS/Aqua	Same as above	HDF/110	NetCDF/296.86
MERRA	Same as above	HDF/312.63	NetCDF/792.27
All 4 data products	Cloud: 4 cores * 2.8GHz/23GB RAM	Same Input/Output as the above.	

Performance Comparison of OPeNDAP

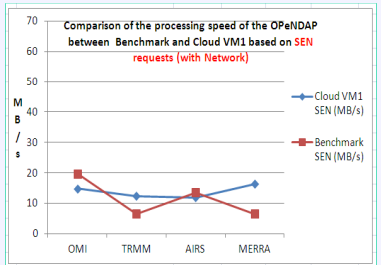
25 granules for each product were pre-staged to the Cloud platform and used to exercise the OPeNDAP server. The OPeNDAP data transfer performance and stability were estimated and compared between the local system and the Cloud platform in different network environment.



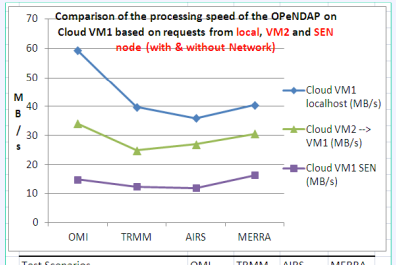
Test Scenarios	OMI	TRMM	AIRS	MERRA
Cloud VM1 localhost (MB/s)	59.2827	39.8284	35.9395	40.4947
Benchmark Localhost (MB/s)	24.1056	17.663	15.926	5.93372
File Size (MB)	68.46	81.25	296.9	792.3



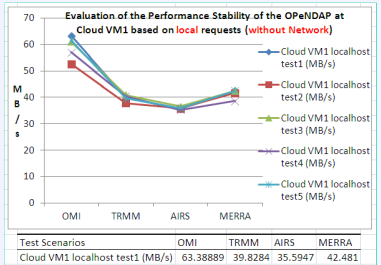
Test Scenarios	OMI	TRMM	AIRS	MERRA
Cloud VM1 CNE (MB/s)	17.2879	19.2993	17.5449	17.4624
Benchmark CNE (MB/s)	11.1136	8.51677	10.233	4.13761
File Size (MB)	68.46	81.25	296.9	792.3



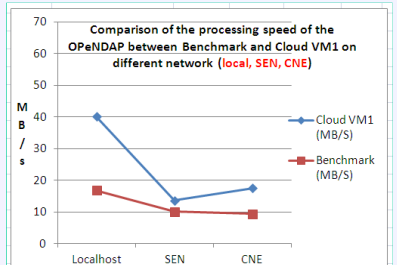
Test Scenarios	OMI	TRMM	AIRS	MERRA
Cloud VM1 SEN (MB/s)	14.8182	12.3857	11.8934	16.3794
Benchmark SEN (MB/s)	19.7291	6.56831	13.5614	6.54931
File Size (MB)	68.46	81.25	296.9	792.3



Test Scenarios	OMI	TRMM	AIRS	MERRA
Cloud VM1 localhost (MB/s)	59.2827	39.8284	35.9395	40.4947
Cloud VM2 --> VM1 (MB/s)	34.0597	24.8471	26.9628	30.5896
Cloud VM1 SEN (MB/s)	14.8182	12.3857	11.8934	16.3794
File Size (MB)	68.46	81.25	296.9	792.3



Test Scenarios	OMI	TRMM	AIRS	MERRA
Cloud VM1 localhost test1 (MB/s)	63.38889	39.8284	35.5947	42.481
Cloud VM1 localhost test2 (MB/s)	52.66154	37.7907	35.8094	41.6327
Cloud VM1 localhost test3 (MB/s)	61.125	40.8291	36.5141	42.5494
Cloud VM1 localhost test4 (MB/s)	57.05	40.4229	35.2985	38.6285
Cloud VM1 localhost test5 (MB/s)	61.67568	39.6341	36.0267	42.5266
File Size (MB)	68.46	81.25	296.9	792.3

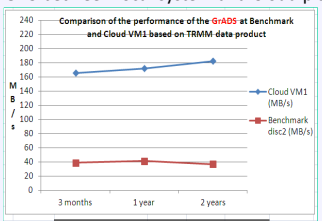


Cloud & Benchmark	Localhost	SEN	CNE
Cloud VM1 (MB/S)	40.1616	13.6019	17.5037
Benchmark (MB/S)	16.7945	10.0649	9.3749

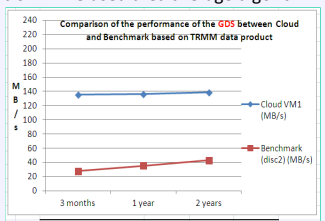
Cloud VM1: Cloud Virtual Machine 1
Cloud VM2: Cloud Virtual Machine 2, same configuration as VM1
CNE: NASA network of 100MB
SEN: NASA network of 1GB
Localhost: without network

Performance Comparison of GDS

TRMM data was used to test and compare the data processing performance of GDS between local system and Cloud platform. We used area average algorithm.



Test scenarios	3 months	1 year	2 years
Cloud VM1 (MB/s)	165.624	172.051	182.4713
Benchmark disc2 (MB/s)	38.9842	41.1685	36.94237
File size (MB)	1588	6439	12878



Test Scenarios	3 months	1 year	2 years
Cloud VM1 (MB/s)	135.3794	135.9873	138.3987
Benchmark (disc2) (MB/s)	27.70895	35.31412	42.92381
Files size (MB)	1588	6439	12878

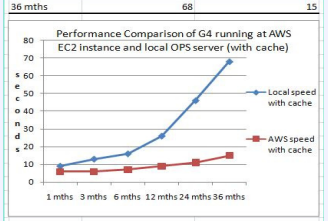
Hardware specifications:

Server	CPU (GHz)	Num. of Cores	RAM (GB)	Notes
Benchmark	2.4 * 4	16	32	AMD 6136
Cloud VM1	2.8 * 1	4	34.2	Intel X5660

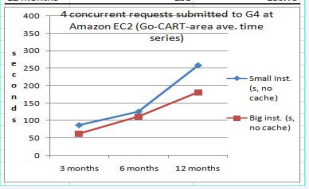
Performance Comparison of Giovanni 4

Giovanni 4 was used to test data analysis in Cloud platform as a case of Data Analysis as a Service and the feasibility of porting long-term legacy-based system to the Cloud platform.

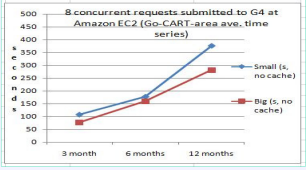
	Local speed with cache	AWS speed with cache
1 mths	9	6
3 mths	13	6
6 mths	16	7
12 mths	26	9
24 mths	46	11
36 mths	68	15



	Small Inst. (s, no cache)	Big inst. (s, no cache)
3 months	86.75	62.25
6 months	125.25	111
12 months	258	180.75



	Small (s, no cache)	Big (s, no cache)
3 month	107.5	76.75
6 months	177.5	158.5
12 months	375.88	280.75



Server	CPU (GHz)	Num. of Cores	RAM (GB)
Benchmark	2 * 4	16	32
EC2 small Inst.	2.8 * 1	4	34.2
EC2 big Inst.	2.8 * 1	8	68.4

* Go-CART: Goddard Chemistry Aerosol Radiation and Transport, a tropospheric aerosol model.

Conclusions and Future Work

A. It is feasible to port legacy satellite data and information services, both simple and complicated ones, to the Cloud platform although the porting process is time-consuming;
B. Cloud computing has better performance, compared with local systems, mostly because Cloud computing has better dedicated servers and better networks;
C. The elasticity of provisioning on-demand multi-cores-oriented resources in the Cloud platform should potentially provide significant advantages to complicated analysis-oriented and computing-intensive applications. This work is ongoing.

Acknowledgements: Thanks the Goddard Information Technology & Communications Directorate for providing the Cloud platform. Authors affiliated with Center for Spatial Information Science and Systems (CSISS), George Mason University have a cooperative agreement with GES DISC (Center Director: **Dr. Liping Di**).